IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Appl. No.	:	10/722,038	Confirmation No.:	6494
Applicant	:	Jeff Peck		
Filed	:	11/24/2003		
TC/A.U.	:	2626		
Examiner	:	Shah, Paras D.		
Docket No.	:	1020.P16469		
Customer No.	:	57035		

Mail Stop AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPEAL BRIEF

This Appeal Brief is in furtherance of the Notice of Appeal filed on September 2, 2009 and the Notice of Panel Decision from Pre-Appeal Brief Review mailed September 29, 2009. The Appeal Brief contains the following sections in the order set forth below:

I. REAL PARTY IN INTEREST

II. RELATED APPEALS AND INTERFERENCES

III. STATUS OF CLAIMS

IV. STATUS OF AMENDMENTS

V. SUMMARY OF THE CLAIMED SUBJECT MATTER

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

VII. ARGUMENT

VIII. CLAIMS APPENDIX

I. REAL PARTY IN INTEREST

The real party in interest in this appeal is Intel Corporation, as the Assignee of record.

II. RELATED APPEALS AND INTERFERENCES

There are no other appeals or interferences that will directly affect, or be affected by, or have a bearing on the decision of the Board in the pending appeal.

III. STATUS OF CLAIMS

Claims originally filed: 1-20

Claims added: None

Claims canceled: 4, 6, 7 and 17-19

Claims withdrawn from consideration: None

Claims allowed: None

Claims objected to: None

Claims rejected: 1-3, 5, 8-16 and 20

Claims on appeal: 1-3, 5, 8-16 and 20

IV. STATUS OF AMENDMENTS

No Amendments have been filed subsequent to the Final Rejection mailed on June 8, 2009.

V. SUMMARY OF THE CLAIMED SUBJECT MATTER

The following is a concise explanation of the subject matter defined in each of the independent claims involved in the appeal. Independent claims 1, 9 and 14 are fully supported, concisely explained, and map to the specification and drawings.

Independent claim 1 maps to the specification and to the drawings as follows.

Claim 1	Specification and Drawings
1. A method, comprising:	[0041] A method and apparatus to perform automatic speech recognition are described. [0012] In one embodiment, system 100 may comprise an Automated Speech Recognition (ASR) system 108. Although ASR system 108 is shown as a separate module for purposes of clarity, it can be appreciated that ASR system 108 may be implemented elsewhere in system 100, such as part of network 104 or call terminal 106, for example.
receiving a plurality of packets with audio information;	[0006] In one embodiment, system 100 may communicate various types of information between the various network nodes. For example, one type of information may comprise audio information. As used herein the term "audio information" may refer to information communicated during a telephone call, such as voice information, silence information, unvoiced information, transient information, and so forth. [0008] In one embodiment, the network nodes may communicate information to

	each other in the form of packets. A packet in this context may refer to a set of information of a limited length, with the length typically represented in terms of bits or bytes. An example of a packet length might be 1000 bytes. [0035] FIG. 3 illustrates a programming logic 300 for an ASR system in accordance with one embodiment. An example of the ASR system may comprise ASR 200. As shown in programming logic 300, a plurality of packets with audio information may be received at block 302.
determining by a voice activity detector whether said audio information represents voice information;	[0013] In one embodiment, ASR 108 may be used to detect voice information from a human user. The voice information may be used by an application system to provide application services. [0027] In one embodiment, ASR 200 may comprise VAD 206. VAD 206 may monitor the incoming stream of information from receiver 202. VAD 206 examines the incoming stream of information on a frame by frame basis to determine the type of information contained within the frame. For example, VAD 206 may be configured to determine whether a frame contains voice information. Once VAD 206 detects voice information, it may perform various predetermined operations, such as send a VAD event message to the application system when speech is detected, stop play when speech is detected (e.g., barge-in) or allow play to continue, record/stream data to the host application only after energy is detected (e.g., voice-activated record/stream) or constantly record/stream, and so forth. The embodiments are not limited in this context.
buffering said audio information in a jitter buffer during said determination; measuring an average packet delay time by said jitter buffer; and	[0031] In one embodiment, jitter buffer 216 may compensate for packets having varying amounts of network latency as they arrive at receiver 202. A transmitter similar

	to transmitter 212 typically sends audio information in sequential packets to receiver 202 via network 104. The packets may take different paths through network 104, or may be randomly delayed along the same path due to changing network conditions. As a result, the sequential packets may arrive at receiver 202 at different times and often out of order. This may affect the temporal pattern of the audio information as it is played out to the listener. Jitter buffer 216 attempts to compensate for the effects of network latency by adding a certain amount of delay to each packet prior to sending them to a voice coder/decoder ("codec"). The added delay gives receiver 202 time to place the packets in the proper sequence, and also to smooth out gaps between packets to maintain the original temporal pattern. The amount of delay added to each packet may vary according to a given jitter buffer delay algorithm. The embodiments are not limited in this context. [0033] In one embodiment, the operations of VAD 206 are performed before or during the operations of jitter buffer 216. This configuration may solve the above-stated problem, as well as others. As a result, the latency normally consumed while the jitter buffer is being filled can be applied to signal processing operations, such as the operations of VAD 206 and any switching to an appropriate endpoint, e.g., to an application system, call terminal for an agent or other intended recipient of the call. In effect, by the time jitter buffer 216 is filled with the active voice information, VAD 206 may have completed its detection operations. The voice information stored in jitter buffer 216 may then be switched to the appropriate endpoint and immediately rendered to the call recipient, without further latency. By performing VAD on an unbuffered stream of audio

	information, it may be possible to save 50-100 milliseconds without degrading performance of ASR 200, for example. It is worthy to note that in a VOP system such as VOP system 100, the contents of pre-buffer 214 may be sent to jitter buffer 216 without inducing additional substantive delay. This approach may be difficult to implement, however, for traditional Time Division Multiplexed (TDM) switched telephony systems.
adding said average packet delay time to each of the plurality of packets prior to sending the plurality of packets to a voice codec;	[0031] In one embodiment, jitter buffer 216 may compensate for packets having varying amounts of network latency as they arrive at receiver 202. A transmitter similar to transmitter 212 typically sends audio information in sequential packets to receiver 202 via network 104. The packets may take different paths through network 104, or may be randomly delayed along the same path due to changing network conditions. As a result, the sequential packets may arrive at receiver 202 at different times and often out of order. This may affect the temporal pattern of the audio information as it is played out to the listener. Jitter buffer 216 attempts to compensate for the effects of network latency by adding a certain amount of delay to each packet prior to sending them to a voice coder/decoder ("codec"). The added delay gives receiver 202 time to place the packets in the proper sequence, and also to smooth out gaps between packets to maintain the original temporal pattern. The amount of delay added to each packet may vary according to a given jitter buffer delay algorithm. The embodiments are not limited in this context.
wherein said determining comprises:	[0027] In one embodiment, ASR 200 may comprise VAD 206. VAD 206 may monitor the incoming stream of information from receiver 202. VAD 206 examines the incoming stream of information on a frame by frame basis to determine the type of information

	contained within the frame. For example, VAD 206 may be configured to determine whether a frame contains voice information. Once VAD 206 detects voice information, it may perform various predetermined operations, such as send a VAD event message to the application system when speech is detected, stop play when speech is detected (e.g., barge-in) or allow play to continue, record/stream data to the host application only after energy is detected (e.g., voice-activated record/stream) or constantly record/stream, and so forth. The embodiments are not limited in this context.
receiving frames of audio information at a voice activity detector; .	[0037] In one embodiment, the determination at block 304 may be made by receiving frames of audio information at a VAD, such as VAD 206. VAD 206 may measure at least one characteristic of the frames. The characteristic may be, for example, an estimate of an energy level for the frame. VAD 206 may determine a start of voice information based on the measurements. VAD 206 may determine an end to the voice information based on the measurements and a delay interval.
measuring at least one characteristic of said frames;	[0028] In one embodiment, estimator 210 of VAD 206 may measure one or more characteristics of the information signal to form one or more frame values. For example, in one embodiment, estimator 210 may estimate energy levels of various samples taken from a frame of information. The energy levels may be measured using the root mean square voltage levels of the signal, for example. Estimator 210 may send the frames values for analysis by VCM 208. [0037] In one embodiment, the determination at block 304 may be made by receiving frames of audio information at a VAD, such as VAD 206. VAD 206 may measure at least one characteristic of the frames. The characteristic may be, for

	example, an estimate of an energy level for the frame. VAD 206 may determine a start of voice information based on the measurements. VAD 206 may determine an end to the voice information based on the measurements and a delay interval.
determining a start of voice information based on said measurements;	[0029] There are numerous ways to estimate the presence of voice activity in a signal using measurements of the energy and/or other attributes of the signal. Energy level estimation, zero-crossing estimation, and echo canceling may be used to assist in estimating the presence of voice activity in a signal. Tone analysis by a tone detection mechanism may be used to assist in estimating the presence of voice activity by ruling out DTMF tones that create false VAD detections. Signal slope analysis, signal mean variance analysis, correlation coefficient analysis, pure spectral analysis, and other methods may also be used to estimate voice activity. The embodiments are not limited in this context. [0037] In one embodiment, the determination at block 304 may be made by receiving frames of audio information at a VAD, such as VAD 206. VAD 206 may measure at least one characteristic of the frames. The characteristic may be, for example, an estimate of an energy level for the frame. VAD 206 may determine a start of voice information based on the measurements. VAD 206 may determine an end to the voice information based on the measurements and a delay interval.
determining an end to said voice information based on said measurements and a delay interval; and	[0037] In one embodiment, the determination at block 304 may be made by receiving frames of audio information at a VAD, such as VAD 206. VAD 206 may measure at least one characteristic of the frames. The characteristic may be, for example, an estimate of an energy level for the frame. VAD 206 may determine a start of voice information based on the measurements. VAD 206 may determine

	an end to the voice information based on the measurements and a delay interval.
adjusting said delay interval to correspond to an average packet delay time.	[0038] In one embodiment, the delay interval may represent a time interval after which VAD 206 determines that voice activity has stopped due to some ending condition, such as termination of a telephone call. Since the operations of VAD 206 may occur prior to buffering by jitter buffer 216, a condition may occur where network latency causes packets to arrive outside the temporal pattern of the voice conversation. This condition may sometimes be referred to as "packet under-run." Consequently, the VAD algorithm implemented by VAD 206 may need to be adjusted to account for packet under-run. Although there are numerous ways to accomplish this, one such adjustment may be to increase the delay time to reduce the potential of artificially detecting an ending condition due to an extended period where packets are not received by receiver 202. This may be accomplished by adjusting the delay interval to correspond to an average packet delay time for the network, such as network 104. The average packet delay time may be predetermined and coded into VAD 206 at start-up. The average packet delay time may also be determined dynamically, and sent to VAD 206 to reflect current network conditions. In the latter case, jitter buffer 216 may measure an average packet delay time, and periodically send the updated average packet delay time to VAD 206.

Independent claim 9 maps to the specification and to the drawings as follows.

Claim 9	Specification and Figures
9. (Currently Amended) A system, comprising:	[0012] In one embodiment, system 100 may comprise an Automated Speech Recognition (ASR) system 108. Although ASR system 108 is shown as a separate module for purposes of clarity, it can be appreciated that ASR system 108 may be implemented elsewhere in system 100, such as part of network 104 or call terminal 106, for example.
an antenna;	[0017] In one embodiment, network 104 may utilize one or more physical communications mediums as previously described. For example, the communications mediums may comprise RF spectrum for a wireless network, such as a cellular or mobile system. In this case, network 104 may further comprise the devices and interfaces to convert the packet signals carried from a wired communications medium to RF signals. Examples of such devices and interfaces may include omni-directional antennas and wireless RF transceivers. The embodiments are not limited in this context. [0023] In one embodiment, ASR 200 may comprise a receiver 202 and a transmitter 212. Receiver 202 and transmitter 212 may be used to receive and transmit information between a network and ASR 200, respectively. An example of a network may comprise network 104. If ASR 200 is implemented as part of a wireless network, receiver 202 and transmitter 212 may be configured with the appropriate hardware and software to communicate RF information, such as an omni-directional antenna, for example. Although receiver 202 and transmitter 212 are shown in FIG. 2 as separate components, it may be appreciated that they may both be combined into a transceiver and still fall within the scope of the embodiments.

a receiver connected to said antenna to receive a frame of information;	[0023] In one embodiment, ASR 200 may comprise a receiver 202 and a transmitter 212. Receiver 202 and transmitter 212 may be used to receive and transmit information between a network and ASR 200, respectively. An example of a network may comprise network 104. If ASR 200 is implemented as part of a wireless network, receiver 202 and transmitter 212 may be configured with the appropriate hardware and software to communicate RF information, such as an omni-directional antenna, for example. Although receiver 202 and transmitter 212 are shown in FIG. 2 as separate components, it may be appreciated that they may both be combined into a transceiver and still fall within the scope of the embodiments. [0027] In one embodiment, ASR 200 may comprise VAD 206. VAD 206 may monitor the incoming stream of information from receiver 202. VAD 206 examines the incoming stream of information on a frame by frame basis to determine the type of information contained within the frame. For example, VAD 206 may be configured to determine whether a frame contains voice information. Once VAD 206 detects voice information, it may perform various predetermined operations, such as send a VAD event message to the application system when speech is detected, stop play when speech is detected (e.g., barge-in) or allow play to continue, record/stream data to the host application only after energy is detected (e.g., voice-activated record/stream) or constantly record/stream, and so forth. The embodiments are not limited in this context.
a voice activity detector to detect voice information in said frame; and	[0027] In one embodiment, ASR 200 may comprise VAD 206. VAD 206 may monitor the incoming stream of information from receiver 202. VAD 206 examines the incoming stream of

	information on a frame by frame basis to determine the type of information contained within the frame. For example, VAD 206 may be configured to determine whether a frame contains voice information. Once VAD 206 detects voice information, it may perform various predetermined operations, such as send a VAD event message to the application system when speech is detected, stop play when speech is detected (e.g., barge-in) or allow play to continue, record/stream data to the host application only after energy is detected (e.g., voice-activated record/stream) or constantly record/stream, and so forth. The embodiments are not limited in this context.
a jitter buffer to buffer said information during said detection by said voice activity detector and to measure an average packet delay time, said jitter buffer to add said average packet delay time to the information prior to sending the information to a voice codec;	[0031] In one embodiment, jitter buffer 216 may compensate for packets having varying amounts of network latency as they arrive at receiver 202. A transmitter similar to transmitter 212 typically sends audio information in sequential packets to receiver 202 via network 104. The packets may take different paths through network 104, or may be randomly delayed along the same path due to changing network conditions. As a result, the sequential packets may arrive at receiver 202 at different times and often out of order. This may affect the temporal pattern of the audio information as it is played out to the listener. Jitter buffer 216 attempts to compensate for the effects of network latency by adding a certain amount of delay to each packet prior to sending them to a voice coder/decoder ("codec"). The added delay gives receiver 202 time to place the packets in the proper sequence, and also to smooth out gaps between packets to maintain the original temporal pattern. The amount of delay added to each packet may vary according to a given jitter buffer delay algorithm. The embodiments are not limited in this context.

	[0033] In one embodiment, the operations of VAD 206 are performed before or during the operations of jitter buffer 216. This configuration may solve the above-stated problem, as well as others. As a result, the latency normally consumed while the jitter buffer is being filled can be applied to signal processing operations, such as the operations of VAD 206 and any switching to an appropriate endpoint, e.g., to an application system, call terminal for an agent or other intended recipient of the call. In effect, by the time jitter buffer 216 is filled with the active voice information, VAD 206 may have completed its detection operations. The voice information stored in jitter buffer 216 may then be switched to the appropriate endpoint and immediately rendered to the call recipient, without further latency. By performing VAD on an unbuffered stream of audio information, it may be possible to save 50-100 milliseconds without degrading performance of ASR 200, for example. It is worthy to note that in a VOP system such as VOP system 100, the contents of pre-buffer 214 may be sent to jitter buffer 216 without inducing additional substantive delay. This approach may be difficult to implement, however, for traditional Time Division Multiplexed (TDM) switched telephony systems.
wherein said voice activity detector receives frames of audio information,	[0037] In one embodiment, the determination at block 304 may be made by receiving frames of audio information at a VAD, such as VAD 206. VAD 206 may measure at least one characteristic of the frames. The characteristic may be, for example, an estimate of an energy level for the frame. VAD 206 may determine a start of voice information based on the measurements. VAD 206 may determine an end to the voice information based on the measurements and a delay interval.

measures at least one characteristic of said frames	[0037] In one embodiment, the determination at block 304 may be made by receiving frames of audio information at a VAD, such as VAD 206. VAD 206 may measure at least one characteristic of the frames. The characteristic may be, for example, an estimate of an energy level for the frame. VAD 206 may determine a start of voice information based on the measurements. VAD 206 may determine an end to the voice information based on the measurements and a delay interval.
determines a start of voice information based on said measurements,	[0037] In one embodiment, the determination at block 304 may be made by receiving frames of audio information at a VAD, such as VAD 206. VAD 206 may measure at least one characteristic of the frames. The characteristic may be, for example, an estimate of an energy level for the frame. VAD 206 may determine a start of voice information based on the measurements. VAD 206 may determine an end to the voice information based on the measurements and a delay interval. .
determines an end to said voice information based on said measurements and a delay interval and	[0037] In one embodiment, the determination at block 304 may be made by receiving frames of audio information at a VAD, such as VAD 206. VAD 206 may measure at least one characteristic of the frames. The characteristic may be, for example, an estimate of an energy level for the frame. VAD 206 may determine a start of voice information based on the measurements. VAD 206 may determine an end to the voice information based on the measurements and a delay interval.
adjusts said delay interval to correspond to said average packet delay time.	[0038] In one embodiment, the delay interval may represent a time interval after which VAD 206 determines that voice activity has stopped due to some ending condition, such as termination of a telephone call. Since the operations of VAD 206 may occur prior to buffering by jitter buffer 216, a condition may occur where network latency causes packets to

	arrive outside the temporal pattern of the voice conversation. This condition may sometimes be referred to as "packet under-run." Consequently, the VAD algorithm implemented by VAD 206 may need to be adjusted to account for packet under-run. Although there are numerous ways to accomplish this, one such adjustment may be to increase the delay time to reduce the potential of artificially detecting an ending condition due to an extended period where packets are not received by receiver 202. This may be accomplished by adjusting the delay interval to correspond to an average packet delay time for the network, such as network 104. The average packet delay time may be predetermined and coded into VAD 206 at start-up. The average packet delay time may also be determined dynamically, and sent to VAD 206 to reflect current network conditions. In the latter case, jitter buffer 216 may measure an average packet delay time, and periodically send the updated average packet delay time to VAD 206.

Independent claim 14 maps to the specification and to the drawings as follows.

Claim 14	Specification and Figures
14. (Currently Amended) An article comprising:	[0012] In one embodiment, system 100 may comprise an Automated Speech Recognition (ASR) system 108. Although ASR system 108 is shown as a separate module for purposes of clarity, it can be appreciated that ASR system 108 may be implemented elsewhere in system 100, such as part of network 104 or call terminal 106, for example. The embodiments are not limited in this context. [0021] FIG. 2 may illustrate an ASR system in accordance with one embodiment. FIG. 2 may illustrate an ASR 200. ASR 200 may be representative of, for example, ASR 108. In one embodiment, ASR 200 may comprise one or more modules or components. For example, in one embodiment ASR 200 may comprise a receiver 202, an echo canceller 204, a VAD 206, and a transmitter 212. VAD 206 may further comprise a Voice Classification Module (VCM) 208 and an estimator 210. Although the embodiment has been described in terms of "modules" to facilitate description, one or more circuits, components, registers, processors, software subroutines, or any combination thereof could be substituted for one, several, or all of the modules.
a computer-readable storage medium;	[0022] The embodiments may be implemented using an architecture that may vary in accordance with any number of factors, such as desired computational rate, power levels, heat tolerances, processing cycle budget, input data rates, output data rates, memory resources, data bus speeds and other performance constraints. For example, one embodiment may be implemented using software executed by a processor. The processor may be a general-purpose or dedicated processor, such as a processor made by Intel® Corporation, for

	example. The software may comprise computer program code segments, programming logic, instructions or data. The software may be stored on a medium accessible by a machine, computer or other processing system. Examples of acceptable mediums may include computer-readable mediums such as read-only memory (ROM), random-access memory (RAM), Programmable ROM (PROM), Erasable PROM (EPROM), magnetic disk, optical disk, and so forth. In one embodiment, the medium may store programming instructions in a compressed and/or encrypted format, as well as instructions that may have to be compiled or installed by an installer before being executed by the processor. In another example, one embodiment may be implemented as dedicated hardware, such as an Application Specific Integrated Circuit (ASIC), Programmable Logic Device (PLD) or Digital Signal Processor (DSP) and accompanying hardware structures. In yet another example, one embodiment may be implemented by any combination of programmed general-purpose computer components and custom hardware components. The embodiments are not limited in this context.
said computer-readable storage medium including stored instructions that, when executed by a processor, result in receiving a plurality of packets with audio information,	[0006] In one embodiment, system 100 may communicate various types of information between the various network nodes. For example, one type of information may comprise audio information. As used herein the term "audio information" may refer to information communicated during a telephone call, such as voice information, silence information, unvoiced information, transient information, and so forth. [0008] In one embodiment, the network nodes may communicate information to each other in the form of packets. A packet in this context may refer to a set of

	information of a limited length, with the length typically represented in terms of bits or bytes. An example of a packet length might be 1000 bytes. [0035] FIG. 3 illustrates a programming logic 300 for an ASR system in accordance with one embodiment. An example of the ASR system may comprise ASR 200. As shown in programming logic 300, a plurality of packets with audio information may be received at block 302.
determining by a voice activity detector whether said audio information represents voice information,	[0013] In one embodiment, ASR 108 may be used to detect voice information from a human user. The voice information may be used by an application system to provide application services. [0027] In one embodiment, ASR 200 may comprise VAD 206. VAD 206 may monitor the incoming stream of information from receiver 202. VAD 206 examines the incoming stream of information on a frame by frame basis to determine the type of information contained within the frame. For example, VAD 206 may be configured to determine whether a frame contains voice information. Once VAD 206 detects voice information, it may perform various predetermined operations, such as send a VAD event message to the application system when speech is detected, stop play when speech is detected (e.g., barge-in) or allow play to continue, record/stream data to the host application only after energy is detected (e.g., voice-activated record/stream) or constantly record/stream, and so forth. The embodiments are not limited in this context.
buffering said audio information in a jitter buffer during said determination,	[0031] In one embodiment, jitter buffer 216 may compensate for packets having varying amounts of network latency as they arrive at receiver 202. A transmitter similar to transmitter 212 typically sends audio information in sequential packets to

	receiver 202 via network 104. The packets may take different paths through network 104, or may be randomly delayed along the same path due to changing network conditions. As a result, the sequential packets may arrive at receiver 202 at different times and often out of order. This may affect the temporal pattern of the audio information as it is played out to the listener. Jitter buffer 216 attempts to compensate for the effects of network latency by adding a certain amount of delay to each packet prior to sending them to a voice coder/decoder ("codec"). The added delay gives receiver 202 time to place the packets in the proper sequence, and also to smooth out gaps between packets to maintain the original temporal pattern. The amount of delay added to each packet may vary according to a given jitter buffer delay algorithm. The embodiments are not limited in this context. [0033] In one embodiment, the operations of VAD 206 are performed before or during the operations of jitter buffer 216. This configuration may solve the above-stated problem, as well as others. As a result, the latency normally consumed while the jitter buffer is being filled can be applied to signal processing operations, such as the operations of VAD 206 and any switching to an appropriate endpoint, e.g., to an application system, call terminal for an agent or other intended recipient of the call. In effect, by the time jitter buffer 216 is filled with the active voice information, VAD 206 may have completed its detection operations. The voice information stored in jitter buffer 216 may then be switched to the appropriate endpoint and immediately rendered to the call recipient, without further latency. By performing VAD on an unbuffered stream of audio information, it may be possible to save 50-100 milliseconds without degrading

	performance of ASR 200, for example. It is worthy to note that in a VOP system such as VOP system 100, the contents of pre-buffer 214 may be sent to jitter buffer 216 without inducing additional substantive delay. This approach may be difficult to implement, however, for traditional Time Division Multiplexed (TDM) switched telephony systems.
measuring an average packet delay time by said jitter buffer, and	[0031] Jitter buffer 216 attempts to compensate for the effects of network latency by adding a certain amount of delay to each packet prior to sending them to a voice coder/decoder ("codec"). The added delay gives receiver 202 time to place the packets in the proper sequence, and also to smooth out gaps between packets to maintain the original temporal pattern. The amount of delay added to each packet may vary according to a given jitter buffer delay algorithm. The embodiments are not limited in this context.
adding said average packet delay time to each of the plurality of packets prior to sending the plurality of packets to a voice codec;	[0031] Jitter buffer 216 attempts to compensate for the effects of network latency by adding a certain amount of delay to each packet prior to sending them to a voice coder/decoder ("codec"). The added delay gives receiver 202 time to place the packets in the proper sequence, and also to smooth out gaps between packets to maintain the original temporal pattern. The amount of delay added to each packet may vary according to a given jitter buffer delay algorithm. The embodiments are not limited in this context.
wherein said determining comprises	[0037] In one embodiment, the determination at block 304 may be made by receiving frames of audio information at a VAD, such as VAD 206. VAD 206 may measure at least one characteristic of the frames. The characteristic may be, for example, an estimate of an energy level for the frame. VAD 206 may determine a start of voice information based on the measurements. VAD 206 may determine an end to the voice information based on

	the measurements and a delay interval.
receiving frames of audio information at a voice activity detector,	[0037] In one embodiment, the determination at block 304 may be made by receiving frames of audio information at a VAD, such as VAD 206. VAD 206 may measure at least one characteristic of the frames. The characteristic may be, for example, an estimate of an energy level for the frame. VAD 206 may determine a start of voice information based on the measurements. VAD 206 may determine an end to the voice information based on the measurements and a delay interval.
measuring at least one characteristic of said frames,	[0037] In one embodiment, the determination at block 304 may be made by receiving frames of audio information at a VAD, such as VAD 206. VAD 206 may measure at least one characteristic of the frames. The characteristic may be, for example, an estimate of an energy level for the frame. VAD 206 may determine a start of voice information based on the measurements. VAD 206 may determine an end to the voice information based on the measurements and a delay interval.
determining a start of voice information based on said measurements,	[0037] In one embodiment, the determination at block 304 may be made by receiving frames of audio information at a VAD, such as VAD 206. VAD 206 may measure at least one characteristic of the frames. The characteristic may be, for example, an estimate of an energy level for the frame. VAD 206 may determine a start of voice information based on the measurements. VAD 206 may determine an end to the voice information based on the measurements and a delay interval.
determining an end to said voice information based on said measurements and a delay interval and	[0037] In one embodiment, the determination at block 304 may be made by receiving frames of audio information at a VAD, such as VAD 206. VAD 206 may measure at least one characteristic of the frames. The characteristic may be, for example, an estimate of an energy level for

	the frame. VAD 206 may determine a start of voice information based on the measurements. VAD 206 may determine an end to the voice information based on the measurements and a delay interval.
adjusting said delay interval to correspond to an average packet delay time.	[0038] In one embodiment, the delay interval may represent a time interval after which VAD 206 determines that voice activity has stopped due to some ending condition, such as termination of a telephone call. Since the operations of VAD 206 may occur prior to buffering by jitter buffer 216, a condition may occur where network latency causes packets to arrive outside the temporal pattern of the voice conversation. This condition may sometimes be referred to as "packet under-run." Consequently, the VAD algorithm implemented by VAD 206 may need to be adjusted to account for packet under-run. Although there are numerous ways to accomplish this, one such adjustment may be to increase the delay time to reduce the potential of artificially detecting an ending condition due to an extended period where packets are not received by receiver 202. This may be accomplished by adjusting the delay interval to correspond to an average packet delay time for the network, such as network 104. The average packet delay time may be predetermined and coded into VAD 206 at start-up. The average packet delay time may also be determined dynamically, and sent to VAD 206 to reflect current network conditions. In the latter case, jitter buffer 216 may measure an average packet delay time, and periodically send the updated average packet delay time to VAD 206.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Whether claims 1, 5, 9, 13 and 14 are unpatentable under 35 U.S.C. § 103(a) over U.S. Patent Publication No. 2004/0073692 to Gentle et al. ("Gentle") in view of U.S. Patent No. 7,346,005 to Dowdal ("Dowdal").

Whether claims 2, 3, 12, 15 and 16 are unpatenable under 35 U.S.C. § 103(a) over Gentle and Dowdal in view of U.S. Patent No. 6,865,162 to Clemm ("Clemm").

Whether claims 8, 10, 11 and 20 are unpatenable under 35 U.S.C. § 103(a) over Gentle and Dowdal in view of U.S. Patent No. 5,920,834 to Sih et al. ("Sih").

VII. ARGUMENT

Rejection Under 35 U.S.C. § 103(a) over U.S. Patent Publication No. 2004/0073692 to Gentle et al. ("Gentle") and U.S. Patent No. 7,346,005 to Dowdal ("Dowdal").

Claims 1, 5 and 14

Applicant respectfully submits that claims 1, 5 and 14 define over the cited references whether taken alone or in combination. For example, independent claim 1 recites the following language, in relevant part:

> A method, comprising:
> receiving a plurality of packets with audio information;
> determining by a voice activity detector whether said audio information represents voice information;
> buffering said audio information in a jitter buffer during said determination;

According to the Office Action, the above recited language is disclosed by the combination of Gentle and Dowdal. In the Final Office Action mailed on June 8, 2009 ("Final Office Action"), the Examiner states that Gentle discloses at paragraphs [0051],

[0051], [0042] and [0061], "buffering said audio information in a jitter buffer during said determination" as recited in claim 1. Applicant respectfully traverses this assertion.

According to the Office Action, the Examiner states that it would be obvious to one of ordinary still in the art that the VAD outputs packets to a second device and that when new audio data is received by the first device, processing by the VAD will occur while the previous packet are being buffered. *See* page 3, Final Office Action. Applicant respectfully submits that this is clearly different than the above recited teaching of claim 1.

Applicant respectfully submits that Gentle fails to teach or suggest all of the limitations contained in claim 1. Paragraph [0051] of Gentle teaches forwarding the results of the jitter to the VAD. However, claim 1 teaches receiving a plurality of packets with audio information, determining… whether said audio information represents voice information and buffering said audio information during said determination. The audio information with the plurality of packets is the same audio information at each step of the limitation. Claim 1 teaches that the audio information is received, that the voice detector determines whether the audio information is voice information and during the determining, the audio information is buffered. Applicant submits that Gentle teaches a serial approach of processing and then buffering the frame of information. Applicant submits that claim 1 is clearly different than the teaching of Gentle.

Applicant respectfully submits that the Examiner has not provided any support in the cited references directed to "buffering said audio information in a jitter buffer during said determination" as recited in independent claim 1. Consequently, Gentle fails to disclose, teach or suggest every element recited in claim 1. Furthermore, Applicant submits that Dowdal fails to remedy the above identified deficiencies of Gentle. For at least these reasons, Applicant submits that claim 1 is patentable over the cited references, whether taken alone or in combination.

In addition, independent claim 14 recites features similar to those recited in claim 1. Therefore, Applicant respectfully submits that claim 14 is not obvious and is patentable over the cited references for reasons analogous to those presented with respect to claim 1. Accordingly, Applicant respectfully requests removal of the obviousness rejection with respect to claim 14.

Furthermore, if an independent claim is non-obvious under 35 U.S.C. § 103, then any claim depending therefrom is non-obvious. *See* MPEP § 2143.03, for example. Therefore, Applicant respectfully requests withdrawal of the obviousness rejection with respect to claim 5 that depends from claim 1 and contains additional features that further distinguishes this claim from the cited references.

Claims 9 and 13

Applicant respectfully submits that claims 9 and 13 define over the cited references whether taken alone or in combination. For example, independent claim 9 recites the following language, in relevant part:

> A system, comprising:
> an antenna;
> a receiver connected to said antenna to receive a frame of information;
> a voice activity detector to detect voice information in said frame; and
> a jitter buffer to buffer said information during said detection by said voice activity detector and to measure an average packet delay time, said jitter buffer to add said average packet delay time to the information prior to sending the information to a voice codec;

According to the Office Action, the above recited language is disclosed by the combination of Gentle and Dowdal. Applicant respectfully traverses this assertion.

The Examiner argues, as above, that the Gentle in paragraphs [0051] and [0052] teach that "when new audio data is received by the first device that processing by the VAD will occur while the previous packets are being buffered." *See* page 9, Final Office Action. Applicant respectfully submits that this is clearly different than the above recited teaching of claim 1.

Applicant respectfully submits that Gentle fails to teach or suggest all of the limitations contained in claim 1. Paragraph [0051] of Gentle teaches forwarding the results of the jitter to the VAD. However, claim 9 teaches a receiver to receive a frame of information, a voice activity detector to detect voice information in said frame, and a jitter buffer to buffer said information during said detection by the voice activity detector. Claim 9 states that the same information is buffered during the detection. Applicant submits that Gentle teaches a serial approach of processing and then buffering the frame of information. Applicant submits that claim 9 is clearly different than the teaching of Gentle.

Applicant respectfully submits that the Examiner has not provided any support in the cited references directed to "a jitter buffer to buffer said information during said detection by said voice activity detector" as recited in independent claim 9. Consequently, Gentle fails to disclose, teach or suggest every element recited in claim 9. Furthermore, Applicant submits that Dowdal fails to remedy the above identified deficiencies of Gentle. For at least these reasons, Applicant submits that claim 9 is patentable over the cited references, whether taken alone or in combination.

Furthermore, if an independent claim is non-obvious under 35 U.S.C. § 103, then any claim depending therefrom is non-obvious. *See* MPEP § 2143.03, for example. Therefore, Applicant respectfully requests withdrawal of the obviousness rejection with respect to claim 13 that depends from claim 9 and contains additional features that further distinguishes this claim from the cited references.

Rejection Under 35 U.S.C. § 103(a) over Gentle and Dowdal in view of US Patent No. 6,865,162 to Clemm ("Clemm").

Claims 2, 3, 12, 15 and 16

Applicant respectfully submits that Clemm fails to remedy the deficiencies of Gentle and Dowdal as discussed above with respect to independent claims 1, 9 and 14. Furthermore, if an independent claim is non-obvious under 35 U.S.C. § 103, then any claim depending therefrom is non-obvious. *See* MPEP § 2143.03, for example. Therefore, Applicant respectfully requests withdrawal of the obviousness rejection with respect to claims 2, 3, 12, 15 and 16 that depend from claims 1, 9 and 14 respectively, and contain additional features that further distinguish these claims from the cited references.

Rejection Under 35 U.S.C. § 103(a) over Gentle and Dowdal in view of US Patent No. 5,920,834 to Sih et al. ("Sih").

Claims 8, 10, 11 and 20

Applicant respectfully submits that Sih fails to remedy the deficiencies of Gentle and Dowdal as discussed above with respect to independent claims 1, 9 and 14. Furthermore, if an independent claim is non-obvious under 35 U.S.C. § 103, then any claim depending therefrom is non-obvious. *See* MPEP § 2143.03, for example. Therefore, Applicant respectfully requests withdrawal of the obviousness rejection with respect to claims 8, 10, 11 and 20 that depend from claims 1, 9 and 14 respectively, and contain additional features that further distinguish these claims from the cited references.

Conclusion

It is believed that claims 1-3, 5, 8-16 and 20 are in allowable form. Accordingly, a timely Notice of Allowance to this effect is earnestly solicited. The Examiner is invited to contact the undersigned at 724-364-3133 to discuss any matter concerning this application. The Office is hereby authorized to charge any additional fees or credit any overpayments under 37 C.F.R. § 1.16 or § 1.17 to the credit card in the previously filed credit card authorization form.

Respectfully submitted,

KACVINSKY LLC

/Rebecca M. Bachner/
Rebecca M. Bachner, Reg. No. 54,865
Under 37 CFR 1.34(a)

Dated: November 2, 2009
KACVINSKY LLC
C/O CPA Global
P.O. Box 52050
Minneapolis, MN 55402

VIII. CLAIMS APPENDIX

1. (Previously Presented) A method, comprising:

receiving a plurality of packets with audio information;

determining by a voice activity detector whether said audio information represents voice information;

buffering said audio information in a jitter buffer during said determination;

measuring an average packet delay time by said jitter buffer; and

adding said average packet delay time to each of the plurality of packets prior to sending the plurality of packets to a voice codec;

wherein said determining comprises:

receiving frames of audio information at a voice activity detector;

measuring at least one characteristic of said frames;

determining a start of voice information based on said measurements;

determining an end to said voice information based on said measurements and a delay interval; and

adjusting said delay interval to correspond to an average packet delay time.

2. (Original) The method of claim 1, further comprising buffering a portion of said audio information in a pre-buffer for a predetermined time interval prior to said determining.

3. (Previously Presented) The method of claim 2, further comprising sending said audio information stored in said pre-buffer and said jitter buffer to an endpoint based on said determination.

4. (Canceled)

5. (Previously Presented) The method of claim 1, wherein said characteristic comprises an estimate of an energy level for said frame.

6. (Canceled)

7. (Canceled)

8. (Original) The method of claim 1, wherein said receiving comprises:

retrieving a frame of audio information from said packets;

receiving an echo cancellation reference signal;

canceling echo from said frame of audio information; and

sending said frame of audio information to a voice activity detector.

9. (Previously Presented) A system, comprising:

an antenna;

a receiver connected to said antenna to receive a frame of information;

a voice activity detector to detect voice information in said frame; and a jitter buffer to buffer said information during said detection by said voice activity detector and to measure an average packet delay time, said jitter buffer to add said average packet delay time to the information prior to sending the information to a voice codec;

wherein said voice activity detector receives frames of audio information, measures at least one characteristic of said frames, determines a start of voice information based on said measurements, determines an end to said voice information based on said measurements and a delay interval and adjusts said delay interval to correspond to said average packet delay time.

10. (Original) The system of claim 9, further comprising an echo canceller connected to said receiver to cancel echo.

11. (Original) The system of claim 10, further comprising a transmitter to provide an echo cancellation reference signal to said echo canceller.

12. (Original) The system of claim 9, further comprising a pre-buffer to store pre-threshold speech during said detection by said voice activity detector.

13. (Original) The system of claim 9, where said voice activity detector further comprises:

an estimator to estimate energy level values; and a voice classification module connected to said estimator to classify information for said frame.

14. (Previously Presented) An article comprising:

a computer-readable storage medium;

said computer-readable storage medium including stored instructions that, when executed by a processor, result in receiving a plurality of packets with audio information, determining by a voice activity detector whether said audio information represents voice information, buffering said audio information in a jitter buffer during said determination, measuring an average packet delay time by said jitter buffer, and adding said average packet delay time to each of the plurality of packets prior to sending the plurality of packets to a voice codec; wherein said determining comprises receiving frames of audio information at a voice activity detector, measuring at least one characteristic of said frames, determining a start of voice information based on said measurements, determining an end to said voice information based on said measurements and a delay interval and adjusting said delay interval to correspond to an average packet delay time.

15. (Original) The article of claim 14, wherein the stored instructions, when executed by a processor, further results in buffering a portion of said audio information in a pre-buffer for a predetermined time interval prior to said determining.

16. (Original) The article of claim 14, wherein the stored instructions, when executed by a processor, further results in sending said audio information stored in said pre-buffer

and said jitter buffer to an endpoint based on said determination.

17. (Canceled)

18. (Canceled)

19. (Canceled)

20. (Original) The article of claim 14, wherein the stored instructions, when executed by a processor, further results in said receiving by retrieving a frame of audio information from said packets, receiving an echo cancellation reference signal, canceling echo from said frame of audio information, and sending said frame of audio information to a voice activity detector.

IX. EVIDENCE APPENDIX

None

X. RELATED PROCEEDINGS APPENDIX

None